

Kathrin Steyer

Wenn der Schwanz mit dem Hund wedelt. Zum linguistischen Erklärungspotenzial der korpusbasierten Kookkurrenzanalyse

1. Einleitung

Eine Herausforderung für die Linguistik ist die Beantwortung der Frage, wie man Wissen über Wörter, vor allem implizites Bedeutungswissen, explizit und damit darstellbar macht. Die Semantiker – ob bekennd kognitiv oder nicht – versuchen es vor allem mit Introspektion, die Psycholinguisten auch mit Tests und Experimenten; in den Wörterbüchern findet man – überwiegend basierend auf der Sprachkompetenz des Lexikografen – Bedeutungssparaphrasen. Sprachkompetenz und in Wörterbüchern gespeichertes sprachliches Wissen stellen zweifellos Orientierungspunkte für derartige Analysen dar. Viel weitreichendere Ergebnisse erzielt man aber, wenn man dabei das empirische Grundprinzip der strikten Korpusbasiertheit verwirklicht. In Textkorpora findet man geronnenes sprachliches Wissen; sie bilden – so Teubert – einen Querschnitt des Diskursuniversums einer Sprachgemeinschaft (vgl. 1999, S. 296). Der heuristische Wert von Korpusdaten für eine Explizierung semantischen Wissens und der Gebrauchsspezifik von Wörtern wird zwar immer mehr erkannt,¹ konsequent korpusbasiertes Arbeiten ist aber nach wie vor eher die Ausnahme in der germanistischen Linguistik. Eine Ursache dafür liegt in der unzureichenden Symbiose zwischen Korpusstechnologie/-recherche einerseits und linguistischer Adaption andererseits. Um in dieser Hinsicht einen qualitativen Fortschritt zu erreichen, braucht man große elektronische Korpora und intelligente Recherchewerkzeuge. Der Philologe muss sich aber vor allem auf neue Denkweisen, auf automatische Analysemethoden und korpusstatistische Prämissen einlassen.

¹ Vgl. dazu ausführlich Haß-Zumkehr (2002).

Im Folgenden soll ein integrativer Ansatz vorgestellt werden, der ein hochkomplexes automatisches Tool, die am IDS von Cyril Belica entwickelte COSMAS-Kollokationsanalyse, mit dem linguistischen Modell der usuellen Kookkurrenz (vgl. Steyer 2000; Belica/Steyer 2002) zu verbinden versucht.² Zunächst als Werkzeug zur Extraktion von Kollokationen und Mehrwortverbindungen gedacht, ist die statistische Kookkurrenzanalyse immer mehr zu einem weit reichenden Konzept geworden, mit dem ein empirischer Zugang zu sprachlichem Wissen möglich wird.³ Der Ansatz wird unter zwei grundsätzlichen Perspektiven diskutiert:

- Es wird erstens gezeigt, dass man sprachliches Wissen generell über usuelle Syntagmen aus Korpora rekonstruieren kann, dass also nicht allein Introspektion zu einer Systematisierung sprachlichen Wissens führt, sondern dies auch und vielleicht sogar vor allem über Kookkurrenzen zu leisten ist.
- Es soll zweitens demonstriert werden, dass mit diesem Ansatz Bedeutungszuschreibungen und Gebrauchsspezifika rekonstruierbar sind, die so nicht im Wörterbuch stehen können und sich bisher weitgehend einer Systematisierung verschließen.

Wir greifen dabei mit dem Kookkurrenzfeld von *Hund* ganz bewusst einen ‘Beispielklassiker’ der Idiomatik auf,⁴ weil auf diese Weise die neuen Möglichkeiten unseres Konzepts besonders deutlich werden.

² Die Basis bilden die Textkorpora des IDS, die mit derzeit über 1,7 Milliarden Textwörtern die weltweit größte Sammlung elektronischer Korpora von geschriebenen deutschsprachigen Texten darstellen. Sie sind online mit dem am IDS entwickelten Werkzeug COSMAS recherchierbar. (<http://corpora.ids-mannheim.de/~cosmas/>)

³ Derzeit wird die Kookkurrenzanalyse vor allem im IDS-Projekt „Wissen über Wörter“, einem Hypertext-Informationssystem zum deutschen Wortschatz, eingesetzt, zum einen zur Extraktion von usuellen Mehrwortverbindungen, zum anderen als lexikologisch-lexikografisches Arbeitsinstrument.

⁴ Vgl. schon Porzig (1934).

2. Zur empirischen Methode der statistischen Kookkurrenzanalyse

2.1 Usuelle Kookkurrenz

Wörter, die in besonderer Weise mit einem anderen Wort verbunden sind, können etwas aussagen über die Bedeutung dieses Wortes und seine Stellung im Kontext, Text bzw. Diskurs. Bei der „idiomatischen Prägung sprachlicher Ausdrücke“ geht es, wie Feilke betont, „um weit mehr als um bloß reproduktive mechanische Routinen des Gedächtnisses, es geht um die Ordnung des sprachlichen Wissens selbst und das ist stets und zuerst eine über Texterfahrung zeichenhaft vermittelte Kompetenz“ (2000, S. 7). So erweisen sich auch Ausdrücke mit gänzlich regulärer Struktur und kompositioneller Bedeutung als hochgradig konventionell und zeichenhaft (*ein Problem behandeln*, *Zähne putzen*) (vgl. ebd.). Diese Formelhaftigkeit lässt sich schwerlich regelhaft allein aus dem Sprachsystem ableiten, die Verbindbarkeit ist arbiträr. Es handelt sich um Gebrauchsnormen bzw. konventionalisierte Kombinationsmodi, um sprachlichen Usus, um **usuelle Kookkurrenz**. Usuelle Kookkurrenzen sind zum einen einzelwortbezogene Kookkurrenzen (im Kern Kollokationen) und zum anderen ‘echte’ Mehrwortverbindungen (u.a. Phraseologismen, Routineformeln, Funktionsverbgefüge usw.).⁵ Kollokationen definieren wir, u.a. aufbauend auf Viehweger, als binäre, usuelle Relationen von Autosemantika. Sie stellen typische Verknüpfungen von Wortschatzelementen dar, die in verschiedenen syntaktischen Strukturen realisiert werden (vgl. 1989, S. 889). Dies ist ein enger Kollokationsbegriff, der nur die binäre Relation zwischen Autosemantika umfasst, dabei aber nicht nur die Kombinationen zwischen unterschiedlichen Wortarten wie Verb–Nomen oder Adjektiv–Nomen, sondern auch solche Kombinationen wie Nomen–Nomen berücksichtigt. Der usuelle Status einer kookkurrenziellen Relation ergibt sich in diesem Konzept aus der statistischen Signifikanz der Wortverbindung, dem Grad der lexikalischen Kohäsion zwischen Wortformen bzw. Types im Korpus. Es geht dabei immer um eine korpusbezogene Usualität.

⁵ Zur linguistischen Begründung und klassifikatorischen Differenzierung dieses Konzepts vgl. Steyer (2000, 2002).

2.2 Hauptaspekte der statistischen Kookkurrenzanalyse

Der Wert der Kookkurrenzanalyse besteht in folgenden Aspekten:

- Sie ermöglicht den Zugang zu Massendaten, wie sie beispielsweise am IDS akquiriert und aufbereitet werden. Ab einer bestimmten Größenordnung von Korpora ist der Mensch mit der Verarbeitung dieser Massendaten überfordert. Er braucht den Rechner, der in einer nicht vergleichbaren Schnelligkeit Häufigkeitsbewertungen und Signifikanzsetzungen vornimmt. Besonders die statistische Kookkurrenzanalyse dient dabei solchen Präferenzbildungen. Der Algorithmus versucht, die innere Struktur der Belegmenge aufzudecken und zu visualisieren, indem er die Kollokatoren des Suchwortes, die Kookkurrenzpartner der Kollokatoren, deren Kookkurrenzpartner usw. sucht und diese hierarchisch anordnet.
- Sie erlaubt kontextsensitive Extraktionen usueller Kookkurrenzen aus großen Korpora. Dabei handelt es sich um ein interaktives und dynamisches Werkzeug, das sich im Gegensatz zu den meisten Tools dieser Art flexibel an unterschiedliche Korpusgegebenheiten anpassen kann und online auf dem neuesten Stand abrufbar ist. Es ermöglicht nicht nur die Erfassung binärer Wortrelationen (Kollokationen), sondern auch usueller phrasaler Muster bis hin zu (idiomatischen) Mehrworteinheiten und ihrer Kontexte.
- Sie eröffnet den empirischen Zugang zu sprachlichem Wissen, da signifikante Kookkurrenzen nicht zufällige Kontextualisierungen darstellen. Auf diesen Aspekt wird im Folgenden ausführlich eingegangen.

2.3 Funktionsweise der COSMAS-Kollokationsanalyse

Den Ausgangspunkt der Analyse bildet die Erfassung des Vorkommens eines Einzelwortes im Korpus. Die entsprechenden Key-Word-in-Context-Belege (KWICs), also die zeilenweise präsentierten Kontexte des Bezugswortes, stellen dann wiederum einen eigenen neuen Sprachausschnitt dar, ein eigenständiges Analysekorpus.

Ausschnitt aus der KWIC-Übersicht für '&Hund' in den IDS-Korpora geschriebener Sprache:

M00/009, Mannheimer Morgen, September 2000

erde aus. Es folgen Kamele, Vögel, **Hunde** . . . Später wird frei model
keine Tür verschlossen, selbst der **Hund** durfte auf Wunsch im Kofferra
eschult", erklärt, der als "harter **Hund**" bekannte Trainer. Vor dem Sa
dem versteckten Lager zurück. Die **Hunde** nahmen dort, wo das Zelt ges
r wissen es nicht. Aber einer, der **Hunde** aufspießt, hat kein reines G
inen größeren Bruch heftete er dem **Hund** ans Halsband, als Dank für di
Birstadt. Besitzer gefährlicher **Hunde** müssen in Birstadt demnächst

R99/JAN, Frankfurter Rundschau, Januar 1999

wenn sie sich zu benehmen wissen. **Hunde**, die hier noch Defizite habe
nen als Insektenvertilger im Haus. **Hunde** dagegen gelten als der Inbeg
nen 14 Jahren auch schon ein alter **Hunde-Herr**, während Minnie immerhi
zur Jagd und wurde von einer Meute **Hunde** unterstützt.
on nichts mit. Künftig stehen sich **Hund** und Passagier gegenüber. Denn
Herr B. einen Maulkorb trägt: sein **Hund** trägt diesen Maulkorb nämlich
den Grabplatz von Kaiser Wilhelms **Hund** kümmert sich heute der Deutsc
mittelbar am Geschehen beteiligten **Hunde** zurück. Gegen die also vom G
[...]

A99/MÄR, St. Galler Tagblatt, März 1999

Freund stellte sich als Bobi, der **Hund**, heraus. Das mit einem Augenz
u+z+ **Hunde**: +z+u
, 8 Monate alt. Ruhiger, sensibler **Hund**. Kaskhu Jagd gerne und brauch
Göttergatte bei der Arbeit und der **Hund** an einem guten Plätzchen unte
sames Hab und Gut, aber auch ihre **Hunde** und Katzen ein Raub der Flam
volle Arbeit für die Gruppe **Jugend+Hund** geehrt und mit einem Geschen
Autobahn: Da wird der Polizei ein **Hund** gemeldet, der auf der A1 umhe
u+z+ **Hund** erschossen +z+u
[...]

Der Rechner versucht nun herauszufinden, ob es in diesem Sprachausschnitt Textwörter gibt, die im Vergleich zum Vorkommen im Gesamtkorpus überproportional häufig miteinander korrelieren – häufiger, als nach dem statistischen Zufallsprinzip zu erwarten gewesen wäre.⁶

⁶ Die Kollokationsanalyse bietet die Möglichkeit der vielfältigen Parametervarianz, die dem Nutzer ganz verschiedene Perspektiven eröffnet und je nach Erkenntnisinteresse äußerst differenzierte Analyseergebnisse hervorbringt. So ist die Kollokationsspanne beliebig einstellbar, können Funktionswörter ein- oder ausgeschlossen werden, und vor allem ist der Grad der Kontextsensitivität der Analyse wählbar. Deshalb ist die Anwendung ohne Reflexion und genaue Bestimmung dessen, was man herausfinden will, nicht effizient möglich.

Man erhält zunächst statistische Listen von Kookkurrenzpartnern zu einem Bezugswort, die schon ohne Einbeziehung der entsprechenden KWICs und Kontexte eine Vielzahl an Informationen liefern, wie an einem Ausschnitt der Liste der Kollokatoren zu *Hund* illustriert werden soll.

Kollokatoren

Analyse-Kontext: 5 Wörter links, 5 Wörter rechts, Autofokus, höchstens 1 Satz

Granularität: grob

Zuverlässigkeit: hoch

Clusterzuordnung: eindeutig

Lemmatisierung: nein

Funktionswörter: ignoriert

Kollokatoren sind sortiert absteigend nach der berechneten Stärke der lexikalischen Kohäsion.

BelegNr	Gamma ⁷	Fokus	Kollokatoren	Häufigkeit ⁸
1+309:	3024	-5;4	Katzen	309
310+142:	1134	-4;5	Katze	142
452+89:	772	2;5	Leine	89
541+58:	659	-3;5	bell	58
599+104:	612	1;1	gekommen	104
703+66:	608	-5;5	Herrchen	66
769+70:	575	-5;5	Halter	70
839+50:	506	-1;-1	gefährlichen	50
889+54:	489	-5;5	Tierheim	54
943+54:	486	-5;5	beißen	54
997+26:	475	-1;-1	bunter <i>bekannt</i> ⁹	26
1023+15:			bunter	15
1038+39:	414	-5;5	bellt	39
1077+31:	358	1;5	Gassi	31
1108+36:	357	-1;-1	gefährliche	36
1144+30:	351	-1;-1	gefährlicher	30
1174+20:	345	-1;-1	streunende	20
1194+34:	340	-1;-1	harter	34
1228+29:	327	-5;5	Leinenzwang	29
1257+20:	322	-1;-1	streunenden	20
1277+27:	311	-4;-2	Halten	27
1304+31:	302	2;2	Katz	31
1335+30:	290	1;5	gebissen	30
1365+39:	289	-5;5	Schwanz	39
1404+27:	243	2;5	Rassen	27
1431+22:	236	-1;-1	Bunte	22
1453+18:	236	-3;5	beißt <i>Mann</i>	18
1471+10:			beißt	10
1481+20:	229	-1;-1	armer	20

⁷ Der Gammawert ist der interne Wert für die ermittelte Stärke der lexikalischen Kohäsion.

⁸ Die Häufigkeitsangabe bezieht sich auf die 10 000 Treffer der statistischen Zufallsauswahl. Bei größeren Belegmengen potenziert sie sich entsprechend.

⁹ Die signifikanten Kookkurrenzpartner der jeweiligen Kollokation sind *Courier* kursiv gesetzt.

1501+30:	223	1;5	spazieren	30
1531+14:	221	-1;-1	wildernde	14
1545+50:	213	-5;5	Besitzer	50
1595+13:	203	-3;5	Frauchen	13
1608+25:	197	-5;5	Hundehalter	25
1633+12:	196	-2;-1	freilaufende	12
1645+21:	191	1;5	angefallen	21
1666+12:	190	-1;-1	freilaufenden	12
1678+21:	187	1;5	gehetzt	21
1699+11:	178	-1;-1	geprügelte	11
1710+13:	178	1;5	angeleint	13
1723+46:	176	-2;-2	Herr	46
1769+14:	173	1;3	Katzenfutter	14
1783+24:	173	1;1	begraben	24
1807+20:	167	-1;-1	arme	20
1827+10:	167	-1;-1	streunender	10
1837+144:	165	-1;-1	zwei	144
1981+6:	159	-1;-1	Streunende	6
1987+20:	147	-4;5	Schafe	20
2007+31:	147	-1;-1	Junge	31
2038+11:	146	-1;-1	bissigen	11
2049+22:	145	-5;3	Pferde	22
2071+8:	144	-1;-1	herrenlose	8
2079+9:	143	-3;-3	Negers	9
2088+8:	141	-3;4	Erziehungskurs	8
2096+12:	141	-1;-1	schlafenden	12
2108+13:	141	-3;5	beißen	13
2121+9:	140	-1;-1	Pawlowschen	9
2130+7:	139	-5;-2	Gefahrenabwehrverordnung	7
2137+10:	136	-5;5	streunen	10
2147+22:	136	-5;5	Pferden	22
2169+14:	135	-5;5	Halsband	14
2183+20:	135	-1;-1	Erwin	20
2203+1:	135	-2;-1	IGP-Vorsitzender	1
2204+14:	132	-1;-1	aggressive	14
2218+13:	131	2;4	Maulkorb	13
2231+8:	129	-1;-1	geprügelter	8
2239+9:	128	1;5	eingeschläfert	9
2248+7:	127	-5;3	Esel	7
2255+17:	125	2;4	Ofen	17
2272+2:	124	2;5	Kleintiere	2
2274+24:	121	-5;5	Tiere	24
2298+14:	121	-3;5	biß	14
2312+18:	120	-1;-1	toten	18
2330+50:	118	-1;-1	kleinen	50
2380+15:	117	-5;3	Spaziergänger	15
2395+8:	115	-4;5	Vögel	8
2403+7:	115	-1;-1	Pawlowscher	7
2410+17:	115	-5;5	Rasse	17
2427+58:	111	-4;-2	Mann	58
2485+26:	108	-1;-1	kleiner	26
2511+11:	108	-4;5	Auslauf	11
2522+9:	108	-3;4	Gebell	9
2531+7:	107	-4;-3	Prokyon	7
2538+7:	107	1;4	anzuleinen	7
2545+12:	105	-1;-1	schlafende	12
2557+10:	102	1;4	vergiftet	10
2567+5:	102	-4;5	Ziegen	5

[...]				
2605+4:	100	-5;4	wedelt	4
[...]				
2749+6:	91	1;2	Kampnagel	6
[...]				
3445+6:	67	1;5	streicheln	6
[...]				
3476+39:	66	-3;-2	Kind	39
3515+2:	66	-5;-2	Haustiere	2
[...]				
4169+7:	54	-5;5	Zucht	7
[...]				
4639+5221:			<i>statistisch unspezifisch</i>	5221 ¹⁰

Diese Listen lassen sich nach ganz unterschiedlichen Kriterien sortieren. Sortiert man sie beispielweise nach Häufigkeit, rücken jene Kollokationen in die oberen Ränge, die eine gewisse Prototypik aufweisen. Sortiert man sie nach Granularität, rücken signifikante nicht-binäre Phrasen an die oberen Stellen, eine Sortierung nach phrasaler Festigkeit also, die einen wichtigen Anhaltspunkt zur Selektion von Idiomen bietet. Wichtige Informationen liefert die Fokusangabe. Mit 'Fokus' wird die typische Stellung des Kollokationspartners zum Bezugswort angegeben. Man kann also bereits anhand dieser Liste Modifikationsanfälligkeiten beurteilen. Fokus -1;-1 bedeutet beispielsweise, dass der Kollokationspartner typischerweise nur unmittelbar vor dem Bezugswort auftritt (*geprügelter Hund*). Diese Kollokationen weisen einen solch engen Grad syntaktischer Gebundenheit auf, dass man von einer ausgeprägten Modifikationsresistenz sprechen kann. Kollokationen mit einer Spanne von -5; 5 (der typische Vorkommensbereich kann in einem Feld von 5 Wörtern vor und nach dem Bezugswort liegen wie bei *Hund-Schwanz*) sind in einem besonders ausgeprägten Maße modifikationsanfällig und damit syntaktisch weniger fest.

¹⁰ Wie unschwer zu erkennen ist, sehen wir uns auch mit Befunden konfrontiert, die man außer Acht lassen muss, wie Eigennamen bzw. Abkürzungen, z.B. *IGP-Vorsitzender*, *Erwin*, *Kampnagel* usw. Überprüft man die KWIC-Zeilen, erkennt man den Zusammenhang:

M98 ... ranzosen zu finden, bedauert **Erwin Hund**, Vorsitzender der IGP ...

T96 +u+ Ein Resümee des Festivals Junge Hunde auf **Kampnagel** ...

2.4 Vom Prototyp zum Stereotyp

Es lässt sich nun überprüfen, ob die Korpusbefunde zum einen überhaupt das abdecken, was man als Wissen über dieses Lexem bereits kodifiziert hat, und ob zum anderen Kontextualisierungen rekonstruierbar sind, die uns das Wörterbuch nicht anbietet. Zum Vergleich sei zunächst der Lexikoneintrag im zehnbändigen Duden (2000) herangezogen:

Hund, der; -[e]s, -e [...] **1.a)** (in vielen Rassen gezüchtetes) kleines bis mittelgroßes Säugetier, das bes. wegen seiner Wachsamkeit u. Anhänglichkeit als Haustier gehalten wird, einen gut ausgebildeten Gehör- und Geruchssinn besitzt u. beißen u. bellen kann [...]

Signifikante Kollokationspartner unserer Liste erscheinen da direkt oder indirekt kompatibel mit den Elementen der Bedeutungsparaphrase im Duden und wohl auch mit unserem Sprecherurteil darüber, was für einen Hund typisch ist. Sie können ohne Kontext interpretiert werden.

Typische Kollokationspartner sind u.a.

<i>Leine</i>	[Hunde werden an der Leine geführt]
<i>bellen</i>	[Hunde bellen]
<i>Herrchen</i>	[Hunde haben einen Besitzer]
<i>Rassen</i>	[Hunde werden in Rassen eingeteilt]
<i>beißen</i>	[Hunde beißen]
<i>Schwanz</i>	[Hunde haben einen Schwanz]
<i>wedelt</i>	[Hunde wedeln mit dem Schwanz]
<i>Gassi</i>	[Mit Hunden geht man nach draußen]
<i>Haustiere</i>	[Hunde gehören zur Gattung der Haustiere]
<i>Zucht</i>	[Hunde werden gezüchtet]
<i>streicheln</i>	[Hunde sind Freunde des Menschen]

Neben diesen als prototypisch zu interpretierenden Kookkurrenzpartnern weist die Liste aber auch eine Vielzahl von Befunden auf, die auf thema-

tisch-diskursive bzw. domänenspezifische Verwendungsweisen hindeuten, bei denen eine **signifikante** Kohäsion ohne Einschränkung auf der Ebene der 'parole' liegt. Sie helfen uns, Aussagen zu so genannten pragmatischen Bedeutungen treffen zu können. So sind Kontextualisierungen von *Hund* ersichtlich, auf die das Wörterbuch nicht eingeht: der Hund als Widerpart des Menschen; der Hund, der als Mittel der Bedrohung dient. Besonders deutlich wird dies, wenn man die Kollokationsanalyse mit dem feinsten Kontextparameter durchführt und somit nicht nur binäre Relationen erhält, sondern weitere statistisch auffällige Kookkurrenzpartner in der Umgebung der Kollokationsrelation, die diese weiter spezifizieren. Die Signifikanz des Miteinandervorkommens bezieht sich nicht nur auf die Relation:

Wenn LEXEM1, dann typischerweise LEXEM2

sondern

Wenn LEXEM1, dann typischerweise LEXEM2 in der Kombination mit LEXEM3-n

oder

Wenn LEXEM1+LEXEM2, dann LEXEM3-n.

Solche phrasalen Festigkeiten können in besonderer Weise darüber Aufschluss geben, welche Konzeptualisierungen nicht nur temporärer Natur sind.

Kollokatoren

Analyse-Kontext: 5 Wörter links, 5 Wörter rechts, Autofokus, höchstens 1 Satz

Granularität: fein

Zuverlässigkeit: analytisch

Clusterzuordnung: eindeutig

Lemmatisierung: nein

Funktionswörter: ignoriert

Kollokationscluster sind sortiert absteigend nach ihrer Granularität

BelegNr	Gamma	Fokus	Kollokatoren	Häufigkeit
8+1:	369	1;5	gebissen kleine angegriffen Mens	1
9+1:	299	-5;5	Besitzer Rasse aggressiven solch	1
10+1:	262	1;5	spazieren gehen Kampfhunde ander	1
[...]				
67+1:	938	2;5	Leine laufen bunte	1
68+1:			Leine Gefährliche müssen	1
69+1:			Leine verboten müssen	1
70+1:	596	-5;5	Herrchen Frauchen angeleinte	1

71+1:			Herrchen Bissige Vierbeiner	1
[...]				
73+1:	486	-5;5	Halter gefährlichen gefährliche	1
74+2:			Halter gefährlicher sind	2
76+1:			Halter gefährliche aggressive	1
77+1:			Halter gefährliche Vierbeiner	1
78+1:			Halter sind eingeschläfert	1
79+1:			Halter sind solchen	1
80+1:			Halter gefährlich Rasse	1
81+1:			Halter ausführen Maulkorb	1
82+1:			Halter Tier müssen	1
83+1:	434	-1;-1	gefährlichen Halten Besitzer	1
84+1:			gefährlichen Halten Leinenzwang	1
85+2:			gefährlichen Halten Führen	2
87+1:			gefährlichen Menschen bekannt	1
88+1:			gefährlichen halten solchen	1
89+1:	424	-5;5	bellt laufen Auto	1
[...]				
93+7:	386	-1;-1	gefährlicher Halten Führen	7
100+2:			gefährlicher Halten Polizeiveror	2
102+1:			gefährlicher Rassen sind	1
103+1:			gefährlicher Haltung Polizeivero	1
104+1:			gefährlicher Zucht verboten	1
105+1:	369	1;5	gebissen angefallen Mann	1
106+1:			gebissen angefallen Spaziergänge	1
107+1:			gebissen angefallen Bein	1
108+1:	368	-1;-1	gefährliche Rassen verboten	1
109+1:	353	-5;-2	Halten Hundehalter Gefahrenabweh	1
110+1:	299	-5;5	Besitzer biß Tier	1
111+1:			Besitzer aggressive Angst	1
112+1:			Besitzer Tier müssen	1
113+1:			Besitzer müssen solchen	1
114+1:	269	-1;-1	armer sind Kind	1
115+1:	266	-4;5	Gassi Hundehalter Spaziergänger	1
116+1:	262	1;5	spazieren Herr Bauschan	1
117+1:			spazieren gehen Auto	1
118+1:			spazieren gehen anderen	1
119+1:	254	-3;-2	Leinenzwang bissige Maulkorb	1
[...]				
125+1:	180	1;5	angefallen zwei Meute	1
126+2:			angefallen zwei großen	2
128+1:			angefallen Gefährliche Menschen	1
[...]				
139+1:	147	-1;-1	freilaufende sind wildernde	1
140+1:			freilaufende sind anderen	1
[...]				
142+1:	136	1;1	sind aggressiv gefährlich	1
143+1:			sind Tiere Menschen	1
[...]				
685+1:	154	-3;-2	Führen Gefahrenabwehrverordnung	1
[...]				
873+1:	92	-1;-1	scharfen aggressiven	1
874+1:			scharfen Angst	1
875+1:			scharfen anschaffen	1
876+1:			scharfen totgebissen	1
956+1:	65	-5;-2	Angst großen	1
957+1:			Angst Auto	1

Konnte man die Thematisierung von *Hund* als Mittel zur Bedrohung von Menschen noch allein anhand der Signifikanzlisten erkennen, erschließen sich pragmatische Zuschreibungen wie beim folgenden Beispiel erst bei der Analyse der Ko- und Kontexte, die man über die KWICs zur jeweiligen Kollokationsrelation und die entsprechenden, variabel erweiterbaren, Volltextbelege erhält. So ist das Nomen *Kind* ein signifikanter Kollokationspartner von *Hund* und wird u.a. in folgender Weise kontextuell eingebettet.

Kind

M95 en hat. Mit **Kind** und Kegel und **Hund** schlendern sie über den Platz
 T90 en Augen des mit Oma, **Kind** und **Hund** erschienenen Publikums streut
 T93 ischen Karriere wegen **Kind** und **Hund** und Mann zurückläßt und Gelde
 T95 en Fotos von Frau und **Kind** und **Hund** und Haus aus der Tasche. Der
 T95 erbringen Mami, Papi, **Kind** und **Hund** heutzutage auf der Straße.

Es ist ein Muster zu erkennen, das auf Grund des signifikanten Auftretens weiterer Kookkurrenzpartner einem sozialen Stereotyp nahe kommt: die Zugehörigkeit des Lexems *Hund* zum Schema 'Familie': Signifikante Kookkurrenzpartner der Kollokation *Hund-Kind* sind u.a. *Kegel (mit Kind und Kegel)*, *Mutter*, *Vater*, *Eltern*, *Familie(n)*, *Frauen*, *Männer*, *Oma*, *Weib*, *Papi*, *Opa*, *Ehefrau*. Erweitert man nun den Kontext der einzelnen Belege, erhält man Ausdifferenzierungen dieses Stereotyps mit unterschiedlichen konnotativen Zuschreibungen und Typisierungen:

Das Lexem 'Hund' erscheint

- als Bestandteil einer typischen Familie:

Eine relativ einfache Sache ist es, den Wecker in die Hand zu nehmen und für den Morgen um eine halbe Stunde vorzustellen, statt auf halb sieben auf sechs Uhr. Wenn er klingelt, düdelt, surrt oder piept – aufstehen! Frau, Mann, **Kinder**, **Hund**, Katze sind zu dieser frühen Stunde weniger bereit, stören zu wollen. Bei Singles erübrigt sich die Sorge, wird aber in den meisten Fällen an der frühen Stunde nichts ändern. (Die Zeit, 2.1.1998, S. 49)

- als Indikator für die 'heile Welt der Familie':

„Frieden für alle“, versichert die CDU Vater, Mutter, **Kind** und **Hund** beim Waldspaziergang in großen Lettern, während nur die kleinen Buchstaben (gegen Krieg, Gewalt und Terror in Europa) an schlimme Dinge anknüpfen. (die tageszeitung, 11.6.1994, S. 5)

– als Mitgestalter einer familiären Atmosphäre:

Zu Fuss auf dem Pass angelangt, war dann alles ganz anders als am Fernsehen. Zaungäste gab es jede Menge, die meisten offenbar Habitues. Jedenfalls rückten viele an mit **Kind** und Kegel, **Hund** und Gartenbestuhlung und manche auch mit Verpflegung. Jung und alt schmorte in der Hitze. Farbtupfer unter die zivilen Zaungäste setzten die bunten Trikots der Freizeitsportler. (Züricher Tagesanzeiger, 13.6.1996, S. 17)

– als Symbol für die ‘Familie als (klein)bürgerliches Korsett’:

Was wollen wir eigentlich? Wohlstand – Beruf – **Hund** oder **Kind** – soziale Betreuung von der Wiege bis zur Bahre, ohne daß in unsere Lebensplanung „etwas Unvorhergesehenes einbricht“? – Ich frage mich manchmal, [...] ob es nicht besser wäre, überhaupt keine Kinder mehr in die Welt zu setzen. Das wäre zumindest sehr viel billiger und gäbe uns mehr Zeit für Freizeitgestaltung und berufliche Karriere. (Frankfurter Allgemeine Zeitung, 1995)

– als Bestandteil der ‘typisch deutschen Familie’:

Es ist die Welt bundesdeutscher Normalniks. Wir sind in Vorstadtsiedlungen und Kinderspielparks. Wir essen Fertiggerichte, spielen Lotto, haben **Kind** und **Hund** und sehen viel fern. (Die Zeit, 20.1.1995, S. 52)

3. Idiomatisches

Anhand von Beispielen aus dem Bereich der Idiomatik soll abschließend gezeigt werden, wie man mit Hilfe der Kookkurrenzanalyse u.a. auch Hinweise zu neuen Lexemen, zu abweichenden Lesarten, zu Invarianzen und zu besonderen Gebrauchsspezifika erhalten kann. Wie schon beschrieben, ist das COSMAS-Werkzeug in der Lage, Kontextmuster zu identifizieren, die in vielen Fällen auf usuelle Mehrwortverbindungen hindeuten.

322+2:	586	1;5	bellens beißen	2
329+4:			bellens Getroffene	4
388+24:	407	-1;-1	bunter bekannt	24
412+1:	392	-4;5	beißen aggressiv	1
413+9:			beißen Letzten	9
427+6:	387	1;1	gekommen sind	6
482+21:			beißt Mann	21
562+6:	225	-5;5	Schwanz wedelt	6
568+5:			Schwanz wedeln	5
574+8:	220	-1;-1	schlafende wecken	8

Hier liegen Kookkurrenzcluster vor: Nicht nur das Verb *bellen* ist ein signifikanter Kollokationspartner von *Hund*, in der Umgebung dieser Kollokation weist auch das Verb *beißen* eine statistisch auffällige Kohäsion mit *Hund-bellen* auf. Ob diese Kookkurrenzcluster tatsächlich ein Ausdruck von Idiomatizität sind, lässt sich nicht ohne die eigene linguistische Kompetenz bzw. die Berücksichtigung der typischen Kontexte bestimmen. So kann man Zweiwortkombinationen nur durch das eigene idiomatische Wissen den Status einer usuellen Mehrwortverbindung zuschreiben (z.B. bei solchen signifikanten adjektivischen Kollokationspartnern wie *hart*, *dick*, *arm*, *scharf*, *toter*). Darüber hinaus gibt es zahlreiche Fälle ‘verdeckter Usualität’ (oft bei nichtidiomatischen Verbindungen), die man erst durch die Kontexte erkennt. Für *Hund* sind u.a. folgende Kollokationspartner zugleich Komponenten einer Mehrwortverbindung: *Katze*, *Ofen*, *bellen*, *beißen*, *begraben*, *gekommen*, *wecken*, *schlafend*, *getroffene*, *bunter*, *armer*, *harter*, *toter*, *geprügelter*, *scharfer*, *gefährlicher*, *dicker*.

3.1 ‘Neue’ Mehrwortverbindungen

Durch die Kollokationsanalyse lassen sich solche Wortverbindungen herausfiltern, die zwar im allgemeinen Sprachgebrauch schon üblich sind, aber noch nicht Eingang in die Wörterbücher gefunden haben.

Beispiel: *toter Hund*

Diese idiomatische Kollokation wird häufig zur Beschreibung der Tatsache verwendet, dass sich etwas oder jemand überlebt hat, „von gestern ist“ und keine Bedeutung mehr besitzt

u+d+ ZITIERT +d+u „Die PDS hat einen festen Platz im bundesdeutschen Parteiensystem. Sie ist ein Sammelbecken der Protestwähler. Die SPD ist kein **toter Hund**. In einigen Monaten wird sie sich erholen. Sie wird dann auf die PDS als stabilen Partner vertrauen können.“ Helmut Holter, PDS-Landesvorsitzender in Mecklenburg-Vorpommern (Berliner Zeitung, 11.10.1999, S. 6)

Kurzum: Sollte all das Reden von stets noch zunehmender postreligiöser Indifferenz nur bloßes Gerede sein? Haben wir überwertige religiöse Fixierungen tatsächlich verabschiedet? Religion sei ein „**toter Hund**“. So hatte Marx in der Mitte des aufklärungsseligen 19. Jahrhunderts dekretiert. Zu Unrecht. (Die Zeit, 26.12.1997, S. 43)

Damit verbunden ist die zweite Verwendungsweise von *toter Hund* als Gleichnis für den achtlosen bzw. verachtenden Umgang mit etwas oder jemandem.

In China dient heute Mao der Legitimation einer Staatsordnung, die für ihn der Inbegriff all dessen gewesen wäre, was er verabscheute. Im Volk wird er verehrt, wie der erste Kaiser und Einiger Chinas, der große, aber grausame Qin Shihuang, mit dem er sich oft verglich. Bei uns im Westen wird er wie ein **toter Hund** traktiert, wenn man vom Interesse für sein Liebesleben absieht. Das wird sich ändern. Denn in unserer hochdifferenzierten, in ihren Subsystemen kreisenden Gesellschaft ist das Bedürfnis nach Utopie unausrottbar. (die tageszeitung, 24.12.1993, S. 3)

In den Texten jener Jahre finden sich Sätze, mit denen er sein politisches, intellektuelles und moralisches Niveau weit unterschritt. Für viele ein Grund, vor seinen Qualitäten die Augen zu schließen. Aber nicht darum wird er heute meist „wie ein **toter Hund**“ behandelt. Strenger noch als für seine politischen Sünden, strafe man ihn für sein Können. Kaum ein Autor der jungen deutschen Literatur, der von ihm nicht gebeutelt wurde. (Frankfurter Allgemeine Zeitung, 1993)

Beispiel: der Schwanz wedelt/wackelt mit dem Hund

Diese Wendung dient als Bezeichnung für die von Sprechern als anachronistisch empfundene Umkehrung eines normalen Prinzips, oft auch verbunden mit einer asymmetrischen Perspektive, z.B.: Minderheiten bestimmen über Mehrheiten, Schwache über Starke, Kleine über Große, Machtlose über Mächtige usw.

Doch Bremens Regierungspolitiker fühlen sich gegenüber dem Wirtschaftswachstum wie der **Schwanz**, der mit dem **Hund** wedeln will. Gewerbeflächen, Straßenneubauten und Qualifizierung von Arbeitskräften [sic!] werden nicht für Daimler geplant, sondern von Daimler gefordert. Und als MBB für den Abtransport der in Bremen produzierten Flügel des neuen Airbus 400 nach einer verlängerten Startbahn verlangte, setzte der Senat im Schnellverfahren und mit Geldgeschenken an die lärmgeplagten Anwohner die Flughafenerweiterung durch. (die tageszeitung, 08.09.1989, S. 9)

„... Unser Koordinatensystem stimmt. Es wird nicht verschoben. Nicht nach links, nicht nach rechts, nicht nach oben, nicht nach unten.“ Wolfgang Schäuble neuer Bundesvorsitzender der CDU „Über kurz oder lang wird der **Schwanz** PDS mit dem **Hund** SPD wedeln.“ (Die Zeit, 13.11.1998, S. 2)

3.2 Invarianten

Mit der statistischen Kookkurrenzanalyse ist es möglich, binnenstrukturelle Validierungen vorzunehmen, z.B. der invarianten semantischen bzw. semantisch-syntaktischen Struktur einer Mehrworteinheit näher zu kommen und somit den obligatorischen Teil zu bestimmen.¹¹

Beispiel: *~allen Hunden gehetzt*

KWIC-Übersicht

B99 önnte. Keiner ist wie er **von allen Hunden gehetzt**, schon gar nicht
 M96 Hooker. 22.45 **Von allen Hunden gehetzt**. 0.30 (SW) Kilom
 M98 **Von allen Hunden gehetzt**. Spionagefilm. 0
 T96 hat und außerdem mit "**Von allen Hunden gehetzt**" eine pferdestar
 S93 Wassern gewaschen, bin **von allen Hunden gehetzt** und habe hinter
 LES/DMS einem Pferd, sie ist **mit allen Hunden gehetzt**. Es hat jetzt au
 THM/AMB Rednertalent, der **mit allen Hunden gehetzt** und in den beden
 U96 aar, das wegen der Beute **von allen Hunden gehetzt** wird, nur um am

Die invariante Formulierung kann aus den jeweils fixen lexikalischen Einheiten und/oder metasprachlichen Markierungen bestehen und sähe für dieses Beispiel folgendermaßen aus:

[PRÄP+allen Hunden gehetzt]¹²

Während *~allen Hunden gehetzt* nicht variiert wird, erscheint neben *von* auch die Präposition *mit*. Invariant ist also hier nicht die lexikalisierte Form, sondern die Kategorie 'Präposition'.

Beispiel: *~hinter dem Ofen hervorlocken*

Die idiomatische Wendung wird fast ausschließlich in der Normalform 'keinen Hund hinter dem Ofen hervorlocken' im Wörterbuch bzw. der Idiomatik-Fachliteratur angegeben. Die KWIC-Zeilen beim Kookkurrenztest zum Basiselement *Ofen* zeigen jedoch, dass dies nicht die invariante Form ist:

¹¹ Zur Problematik der ('invarianten') Normalform als subjektive Abstraktion vgl. vor allem Dobrovol'skij (1993) und Burger (1998), auch Fleischer (1997).

¹² In diesem Fall ist sogar die flektierte Form invariant; im Regelfall werden die flektierten Formen auf ihre Grundform zurückgeführt.

T92 gentlich keinen Hund hinter dem Ofen hervorlocken. Und trotzdem b
 U97 ers wohl keinen Hund hinter dem Ofen hervorlocken", dürfte die er
 R97 en damit keinen Hund hinter dem Ofen hervorlocken. Die intellektu
 E00 de keinen müden Hund hinter dem Ofen hervorlocken. Dazu braucht e
 U98 keinen Menschen mehr hinter dem Ofen hervorlocken. Auf den Plakat
 B98 ßlich niemanden mehr hinter dem Ofen hervorlocken. Aus der losen
 M99 kein Nachfolger mehr hinter dem Ofen hervorlocken.
 B98 g kaum noch jemanden hinter dem Ofen hervorlocken. Der erneute Ni
 T92 ann so leicht nichts hinter dem Ofen hervorlocken. Und doch: stat
 T99 A noch weniger Leute hinter dem Ofen hervorlocken als eine weiter
 Z97 été!" noch irgendwen hinter dem Ofen hervorlocken zu können. "Oc
 M00 en locken nur wenige hinter dem Ofen hervor. Die Formel "global d

Invariant ist hier nur *hinter dem Ofen hervorlocken*. Das Lexem *Hund* stellt zwar das prototypische Basiselement dar; es ist aber austauschbar, z.B. durch 'Mensch'. Die Negation ist gleichfalls ein kategoriales Basiselement, das allerdings nicht in allen Fällen durch die Quantitätsverneinung *kein* repräsentiert ist.¹³

Die invariante Form lautet demzufolge:

[NEG+ (fak Nomen)+hinter dem Ofen hervorlocken]¹⁴

Pilotuntersuchungen im Rahmen des IDS-Projektes „Wissen über Wörter“ haben dabei gezeigt, wie sehr die eigene Intuition gerade in Bezug auf solche invarianten Kerne täuschen kann.

3.3 Semantische Informationen und Gebrauchsspezifik

Die Kookkurrenz- und Kontextanalyse der Belege bietet schließlich eine Vielzahl von Informationen, die man durch Introspektion oder durch die Analyse einzelner Texte in keinem Fall gewinnen kann. Ein typisches Beispiel dafür ist die relativ 'neue' Kollokation *harter Hund*. In den IDS-Korpora ist sie erst seit 1991 belegt; bei den Printwörterbüchern hat sie bisher nur der zehnbändige Duden (2000) aufgenommen:

„[...] ; der neue Trainer gilt als harter H. (ugs.; als Anhänger harter Trainingsmethoden); [...]“

¹³ Vgl. zu den unterschiedlichen Negationstypen u.a. Stickel (1970), Hentschel (1998)

¹⁴ Im Korpus gleichfalls enthaltene Formulierungen ohne explizite Negationsmarker sind okkasionell.

Korpusbasiert lässt sich folgende Paraphrase formulieren:

wertende Bezeichnung für einen Mann (oft in einer Führungsposition), der konsequent, u.U. ohne Rücksicht auf seine Umgebung, seine Ziele durchsetzt

Harter Hund wird dabei häufig in Verbindung mit Nomina wie *Härte, Angst, Respekt, Disziplin, Ordnung* und *Durchsetzungsvermögen* gebraucht. Gleichzeitig werden typische Eigenschaften eines solchen Mannes¹⁵ durch die Kontrastierung mit Eigenschaften manifest gemacht, die der Sprecher für gegenständig hält, wie *Weichheit, Feinsinnigkeit, menschliche Sensibilität* oder *Nachgiebigkeit*.

„Natürlich kann man Erfolg kaufen“, meint der wortkarge Trainer des Fußball-Regionalligisten Tennis Borussia Berlin. Zum Philosophieren wurde der als „**harter Hund**“ verschriene Sportlehrer („Ich geh' gradeaus durch drei Wände“) auch nicht geholt. (die tageszeitung., 29.1.1998, S. 23)

Raschen Einordnungen entzieht sich der neue Bezirksleiter, dessen Leidenschaft das Lesen ist und dem nicht gerade der Ruf vorausseilt, „ein **harter Hund**“ zu sein. (Frankfurter Rundschau, 26.11.1998, S. 5)

Diese Kontextualisierungen werden z.B. anhand paradigmatischer Relationen deutlich, die aus der eigenen Sprachkompetenz nicht ohne weiteres ableitbar wären. Erkennbar sind z.B. solche Antonyme wie *Softie, Kumpeltyp, Weichspüler* aus expliziten Negationsmustern in den KWIC-Zeilen [Er ist ein harter Hund, **kein** Softie].

Der Gebrauch dieser Wendung ist im Korpus an bestimmte Domänen gebunden; dort wiederum werden spezielle Berufsgruppen zu präferierten Referenzobjekten. Der frequenteste Gebrauch ist in der Domäne 'Sport' zu verzeichnen und hier fast ausschließlich für Vertreter der Berufsgruppe 'Trainer'. In der Domäne 'Wirtschaft' werden vorzugsweise Manager, besonders Sanierer maroder Wirtschaftsunternehmen, in dieser Art und Weise bezeichnet. In der Domäne 'Politik, Staat und Recht' betrifft die Zuschreibung insbesondere Anwälte und Vertreter der öffentlichen Ordnung.

Korpusbasiert lassen sich typische Verwendungsmuster systematisieren wie:

¹⁵ 'Harter Hund' als Bezeichnung für eine Frau ist im Korpus nicht belegt.

als harter Hund gelten, als harter Hund verschrien sein, den Ruf/das Image als harter Hund haben, jemandem eilt der Ruf/das Image eines harten Hundes voraus, den harten Hund raushängen lassen

Besonders aufschlussreich sind bei diesem Beispiel die mit der Verwendung der Kollokation verbundenen Sprechereinstellungen und -bewertungen. Analysiert man die Korpusbelege zu *harter Hund*, so lassen sich eindeutige Zuordnungen zu einem negativen bzw. positiven Konnotationsfeld nur selten vornehmen. Vielmehr changieren die Einstellungen der Sprecher in einem sehr differenzierten Bewertungsspektrum, das durch seine Vagheit den Hörern auch einen größeren Interpretationsspielraum ermöglicht. In den meisten Fällen werden die Zuschreibungen in weitere Bewertungskontexte eingebettet. Das heißt, Sprecher positionieren sich zu ihrer eigenen Zuschreibung. Ein impliziter Wertekonsens scheint die Sprecher augenscheinlich dazu zu veranlassen, ihre eigene Bewertung zu thematisieren, zu relativieren bzw. zu verteidigen. Dabei gibt es abgestufte argumentative Varianten.

- Explizit negativ: Sprecher bewerten die Zuschreibung, ein *harter Hund* zu sein.

In der deutschen Managerelite weht der Wind der Veränderung. Die patriarchalisch geprägte, von Nationalsozialismus und Wiederaufbaustolz bestimmte Managergeneration, in der Ingenieursdenken, Produktstolz und ein selbstgefälliges Baden im Wohlgefühl wirtschaftspolitischer Gesamtverantwortung zur Grundausrüstung gehörten, räumt seit einigen Jahren die Chefessel für „**harte Hunde**“. So jedenfalls beschreibt Norddeutschlands IG-Metall-Chef Frank Teichmüller die Newcomer, denen Traditionen, Vaterland und das liebevolle Streicheln einer Maschine made in Germany piegegal sind. Jürgen Schrempps erster Blick morgens im Büro gilt denn auch dem Reuter-Monitor, der online jede Zuckung der Daimler-Aktie registriert. Wo früher Marktstrategen, Visionäre und Techniker die Vorstandsflure beherrschten, sitzen, so erläutert Hamburgs IG-Metall-Boß Klaus Mehrens, „heute oft Finanzmanager und Controller“ am Ruder. (die tageszeitung, 6.3.1996, S. 3)

- Explizit positiv: In anderen Sequenzen wird die negative Bewertung überhaupt nicht explizit gemacht und ausschließlich im positiven Sinne argumentiert.

Corts habe den Beweis dafür antreten müssen, daß auch er ein **harter Hund** sein könne. So erreichte ihn denn auch ein dickes Lob von seiner Konkurrenz für seine „klare Haltung“. Erika Steinbach: „Ein fähiger Mann.“ Ordnung

muß schließlich das Credo eines Ordnungsdezernenten sein. (die tageszeitung, 17.6.1997, S. 5)

Roland Kochs Ideen zur Sozialhilfe sind so ein Fall. Sein Vorstoß passt einfach zu gut zu dem Image als **harter Hund**, das der hessische Ministerpräsident so hingebungsvoll kultiviert. (Süddeutsche Zeitung, 7.8.2001, S. 4; bisher noch nicht in den IDS-Korpora enthalten)

- Positive Einbettung: Ein Sprecher sagt, jemand hat zwar diese Eigenschaft, die im allgemeinen Verständnis negativ zu bewerten ist, aber unter bestimmten Umständen sind solche Verhaltensweisen auch positiv zu werten. Mit dieser adversativen Konstruktion wird zwar das negative Konnotationspotenzial eingeräumt, aber es erfährt eine teilweise Umwertung.

Ein ehemaliger Mitarbeiter der Vereinsbankgruppe urteilt gar: „Wenn es sein muss, geht der über Leichen.“ Unfair, gar brutal, nein diesen Vorwurf will ein Mitglied des Vorstandes nicht gelten lassen. „Harter, sehr **harter Hund**, einverstanden. Aber immer kollegial.“ (Die Zeit, 19.11.1998, S. 37)

Diese adversative argumentative Einbettung ist bereits bei den Erstbelegen festzustellen und stellt den typischen Fall dar. Deshalb kann man hier durchaus von einer diskursiven Stereotypisierung sprechen. Zumeist drückt ein Sprecher neben allen Distanzierungssignalen auch eine gewisse Anerkennung für ein derartiges soziales Verhalten aus. Durch die Analyse des Kookkurrenzfeldes von *harter Hund* und der damit verbundenen Kontextmuster lässt sich also eine sehr viel differenziertere Beschreibung der Bedeutung und der Gebrauchsspezifika dieses Idioms vornehmen, als das mit der eigenen Kompetenz bzw. auf der Basis einzelner Texte möglich ist.

4. Schlussbemerkung

Es sollte deutlich geworden sein, dass die Kookkurrenzanalyse ein heuristisches Instrument zur linguistischen Analyse darstellt, ein Instrument, das der Methode der reinen Introspektion weit überlegen ist. Es sollte auch deutlich geworden sein, dass linguistische Kompetenz dabei in keiner Phase durch den Computer zu ersetzen ist. Die Entscheidung, wie mit welcher Korpusinformation umzugehen ist und welche Schlussfolgerungen zu ziehen sind, bleibt dem Menschen vorbehalten. Aber mit Hilfe von Korpusdaten und automatischen Analysewerkzeugen wird es dem Linguisten ermöglicht, sich

in der Welt der Massendaten und der Informationsflut zu orientieren und für ihn relevante sprachliche Informationen zu extrahieren. Die Kookkurrenzanalyse hilft schließlich, den Blick für Phänomene des sprachlichen Usus zu schärfen, vor allem für Erscheinungen, die sich bisher einer Systematisierung verschließen. Wenn man sich mit statistischen Listen beschäftigt, mit KWICs und massenhaften Belegreihen, erkennt man Rekurrentes, Muster und geordnete Strukturen. Man findet neue Hypothesen und Prämissen. Man kommt zu Fragen, die man sonst gar nicht gestellt hätte. Besonders für den Bereich der korpusbasierten Pragmatik scheinen diese Methoden einen wirklich neuartigen Weg zur empirischen Verifizierung pragma-kommunikativer Modelle und Theoreme zu eröffnen. Wir sind erst am Anfang.

5. Literatur

- Belica, Cyril/Steyer, Kathrin (2002): Die COSMAS-Kollokationsanalyse – statistisches Modell, Funktionsweise und Interpretationsspielräume. (<http://www.ids-mannheim.de/kt> (in Vorber.)).
- Burger, Harald (1998): Phraseologie. Eine Einführung am Beispiel des Deutschen. Berlin. (= Grundlagen der Germanistik 36).
- Dobrovol'skij, Dmitrij (1993): Datenbank deutscher Idiome. Aufbauprinzipien und Einsatzmöglichkeiten. In: Földes, Csaba (Hg.): Germanistik und Deutschlehrrausbildung. Szeged/Wien. S. 51-67.
- Duden (2000): Das große Wörterbuch der deutschen Sprache. 10 Bde. Mannheim.
- Feilke, Helmuth (2000): Textroutine, Textsemantik und sprachliches Wissen. Bielefeld. (Vortragsmanuskript).
- Fleischer, Wolfgang (1997): Phraseologie der deutschen Gegenwartssprache. 2., durchges. u. erg. Aufl. Tübingen.
- Haß-Zumkehr, Ulrike (2002): Das Wort in der Korpuslinguistik. Chancen und Probleme empirischer Lexikologie. In: Agel, Vilmos/Gardt, Andreas/Haß-Zumkehr, Ulrike/Roelcke, Thorsten (Hg.): Das Wort ... Festschrift für Oskar Reichmann zum 65. Geburtstag. Tübingen. (ersch.)
- Hentschel, Elke (1998): Negation und Interrogation. Studien zur Universalität ihrer Funktionen. Tübingen (= Reihe Germanistische Linguistik 195).
- Porzig, Walter (1934): Wesenhafte Bedeutungsbeziehungen. In: Beiträge zur Geschichte der deutschen Sprache und Literatur 58, S. 70-97.

- Steyer, Kathrin (2000): Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikografische Möglichkeiten. In: Deutsche Sprache 2/00, S. 101-125.
- Steyer, Kathrin (2002): Idiomatik hypermedial. Zur Repräsentation von Wortverbindungen im Informationssystem „Wissen über Wörter“. In: EUROPHRAS 2000 (ersch.).
- Stickel, Gerhard (1970): Untersuchungen zur Negation im heutigen Deutsch. Braunschweig. (= Schriften zur Linguistik 1).
- Teubert, Wolfgang (1999): Korpuslinguistik und Lexikographie. In: Deutsche Sprache 4/1999, S. 292-313.
- Viehweger, Dieter (1989): Probleme der Beschreibung semantischer Vereinbarkeitsrelationen im allgemeinen einsprachigen Wörterbuch. In: Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (Hg.): Wörterbücher. Ein internationales Handbuch zur Lexikographie. Berlin/New York (= HSK 5.1). S. 888-893.